Self-reported Work Disability in the US and The Netherlands

On-line Appendices

Appendix 1: Vignette descriptions used in both the U.S. and The Netherlands

This appendix presents the 15 vignettes on work disability used in the analysis. The numbers of the vignettes correspond to those in the descriptive statistics in Table 2 of the paper.

Vignettes for Affect

1. [Henriette] generally enjoys her work. She gets depressed every 3 weeks for a day or two and loses interest in what she usually enjoys but is able to carry on with her day-to-day activities on the job.

2. [Jim] enjoys work very much. He feels that he is doing a very good job and is optimistic about the future.

3. [Tamara] has mood swings on the job. When she gets depressed, everything she does at work is an effort for her and she no longer enjoys her usual activities at work. These mood swings are not predictable and occur two or three times during a month.

4. [Eva] feels worried all the time. She gets depressed once a week at work for a couple of days in a row, thinking about what could go wrong and that her boss will disapprove of her condition. But she is able to come out of this mood if she concentrates on something else.

5. [Roberta] feels depressed most of the time. She weeps frequently at work and feels hopeless about the future. She feels that she has become a burden to her co-workers and that she would be better dead.

Vignettes for Pain

1. [Katie] occasionally feels back pain at work, but this has not happened for the last several months now. If she feels back pain, it typically lasts only for a few days.

2. [Catherine] suffers from back pain that causes stiffness in her back especially at work but is relieved with low doses of medication. She does not have any pains other than this generalized discomfort.

3. [Yvonne] has almost constant pain in her back and this sometimes prevents her from doing her work.

4. [Jim] has back pain that makes changes in body position while he is working very uncomfortable. He is unable to stand or sit for more than half an hour. Medicines

decrease the pain a little, but it is there all the time and interferes with his ability to carry out even day to day tasks at work.

5. [Mark] has pain in his back and legs, and the pain is present almost all the time. It gets worse while he is working. Although medication helps, he feels uncomfortable when moving around , holding and lifting things at work

Vignettes for CVD

1. [Trish] is very active and fit. She takes aerobic classes 3 times a week. Her job is not physically demanding, but sometimes a little stressful.

2. [Norbert] has had heart problems in the past and he has been told to watch his cholesterol level. Sometimes if he feels stressed at work he feels pain in his chest and occasionally in his arms.

3. [Paul]'s family has a history of heart problems. His father died of a heart attack when Paul was still very young. The doctors have told Paul that he is at severe risk of having a serious heart attack himself and that he should avoid strenuous physical activity or stress. His work is sedentary, but he frequently has to meet strict deadlines, which adds considerable pressure to his job. He sometimes feels severe pain in chest and arms, and suffers from dizziness, fainting, sweating, nausea or shortness of breath

4. [Tom] has been diagnosed with high blood pressure. His blood pressure goes up quickly if he feels under stress. Tom does not exercise much and is overweight. His job is not physically demanding, but sometimes it can be hectic. He does not get along with his boss very well.

5. [Dan] has undergone triple bypass heart surgery. He is a heavy smoker and still experiences severe chest pain sometimes. His job does not involve heavy physical demands, but sometimes at work he experiences dizzy spells and chest pain.

Appendix 2: Parameter Estimates Benchmark Model

This appendix presents all parameter estimates and standard errors of the benchmark model, including the parameters in Table 3 of the paper but also the (auxiliary) parameters not included in Table 3.

Work disability (β)

	US parameters		Interaction with dummy NL	
	par.	s.e.	par.	s.e.
constant	-7.911	0.356*	1.939	0.599*
educ. interm.	-3.776	0.491*	3.123	0.960*
education high	-6.568	0.623*	3.947	1.051*
age/100	151.656	27.526*	-103.462	31.143*
(age/100)^2	-113.753	21.980*	75.421	26.062*
female	0.622	0.417	0.404	0.814
hypertension	2.548	0.427*	-1.793	0.942
diabetes	2.650	0.633*	1.945	1.727
cancer	2.970	0.637*	-0.424	1.591
disease of lung	6.175	0.770*	0.028	1.523
heart problem	5.876	0.672*	2.636	1.434
emotional	4.673	0.644*	2.368	1.156*
pain	10.921	0.515*	4.243	0.927*

First threshold (γ^1)

	US parameters		Interaction with dummy NL	
	par.	s.e.	par.	s.e.
constant	0		-3.046	0.269*
educ. interm.	-0.466	0.395	0.344	0.442
education high	-1.029	0.461*	1.388	0.508*
age/100	50.195	12.316*	-51.807	12.609*
(age/100)^2	-49.727	10.654*	50.646	10.959*
female	1.047	0.323*	-1.070	0.358*
hypertension	0.413	0.335	-0.320	0.391
diabetes	-1.582	0.551*	0.681	0.760
cancer	0.229	0.502	-0.226	0.831
disease of lung	-0.008	0.651	-0.035	0.797
heart problem	-0.197	0.618	0.391	0.710
emotional	-1.990	0.527*	1.140	0.594
pain	-0.644	0.331	0.949	0.397*

Second threshold(γ^2)

	US parameters		Interaction with dummy NL	
	par.	s.e.	par.	s.e.
constant	1.883	0.045*	0.333	0.024*
educ. interm.	0.073	0.041	-0.065	0.043
education high	0.125	0.050*	-0.129	0.052*
age/100	0.310	1.497	-0.515	1.512
(age/100)^2	-0.215	1.311	0.377	1.328
female	-0.094	0.032*	0.113	0.034*
hypertension	-0.036	0.037	0.076	0.039
diabetes	-0.068	0.057	0.152	0.064*
cancer	-0.061	0.060	0.056	0.067
disease of lung	-0.122	0.062*	0.146	0.069*
heart problem	-0.003	0.084	0.002	0.087
emotional	0.032	0.051	-0.022	0.054
pain	0.081	0.035*	-0.096	0.037*

Third threshold (γ^3)

	US parameters		Interaction with dummy NL	
	par.	s.e.	par.	s.e.
constant	1.783	0.046*	0.040	0.027
educ. interm.	-0.111	0.050*	0.125	0.052*
education high	-0.023	0.059	0.049	0.061
age/100	0.965	1.822	-1.301	1.841
(age/100)^2	-0.599	1.600	0.968	1.623
female	-0.116	0.039*	0.119	0.041*
hypertension	-0.029	0.043	0.071	0.046
diabetes	0.096	0.058	-0.157	0.070*
cancer	-0.032	0.077	0.004	0.086
disease of lung	0.034	0.080	-0.085	0.089
heart problem	-0.125	0.088	0.127	0.093
emotional	0.027	0.052	-0.018	0.057
pain	-0.075	0.041	0.042	0.044

Fourth threshold (γ^4)

	US parameters		Interaction with dummy NL	
	par.	s.e.	par.	s.e.
constant	2.119	0.044*	-0.058	0.024*
educ. interm.	0.027	0.049	0.012	0.052
education high	0.049	0.059	0.034	0.061
age/100	4.210	1.571*	-3.995	1.599*
(age/100)^2	-3.312	1.313*	3.317	1.343*
female	0.051	0.035	-0.029	0.039
hypertension	-0.014	0.039	-0.001	0.044
diabetes	-0.071	0.052	0.084	0.067
cancer	-0.042	0.058	0.116	0.067
disease of lung	-0.119	0.074	0.133	0.082
heart problem	0.194	0.065*	-0.233	0.072*
emotional	-0.004	0.051	-0.005	0.057
pain	-0.019	0.038	-0.066	0.043

Vignette Evaluations $(\theta_l, l = 1,...,15; \theta_1^f, \theta_2^f, \theta_3^f)$

	Affect		Pain		CVD	
	par.	s.e.	par.	s.e.	par.	s.e.
dum vig1	-1.056	0.232*	0.455	0.235	-10.636	0.472*
dum vig2	-13.783	0.584*	4.159	0.286*	5.587	0.317*
dum vig3	5.703	0.322*	14.672	0.621*	11.897	0.516*
dum vig4	4.028	0.279*	14.716	0.623*	3.584	0.264*
dum vig5	16.037	0.661*	12.958	0.561*	8.328	0.399*
vign. female	-0.092	0.061	-0.364	0.061*	-0.919	0.069*

Standard Deviations of Error Terms

	par.	s.e.
Work disability (σ_r)	10	(normalized)
Thresholds (σ_u)	3.629	0.148*
Vignette evaluations (σ)	5.517	0.213*
Self-report 2 point (σ_2)	-4.894	0.671*
Self-report 5 point (σ_5)	5.981	0.505*

Transformation from 5-point to 2-point scale

	par.	s.e.
λ	0.758	0.051*

Appendix 3: General Model with Multiple Domains of Work-Releated Health

Since the three variants with the response scales based on vignettes for just one domain lead to substantially different conclusions about response scale differences between the US and the Netherlands (see Table 5 in the paper), we also formulate a "general model" that does not impose that response scales are the same across the three domains. Separate models for the three domains are not an option, since we only observe one self-report on "overall" work disability. The model we introduce models observed overall work disability as the worst of the (unobserved) outcomes in the separate work disability domains. In other words, in this model it is assumed that true work limitations are the maximum of work limitations in several domains. The domains are Affect, Pain, CVD, and "Other". For the former three we have vignettes, which can be used to correct scale differences across the two countries, but we have no vignettes for "Other". For "Other" we will assume in the simulations that the response scale is a weighted average of the response scales for affect, pain and CVD, with weights determined by the contribution of each of these three domains to overall work disability (as implied by the model estimates).

To identify this general model, we make some plausible exclusion restrictions concerning work disability in a specific domain. For the domains Affect, Pain and CVD, we assume that only respondents who have the corresponding health problem can report work disability due to a problem in that domain. For example, the equation for work disability in the affect domain only applies to respondents reporting that the doctor has ever told them that they have an emotional health problem.

To be precise, let the domains determining work related health be given by *d=1,2, 3* and *4—affect, pain, heart problems*, and *other*. For the first three (*d=1,2,3*), we assume that only those who report a health condition in that domain can suffer from a work disability in that domain. Such an assumption is not made for *other* (*d=4*), to avoid assuming a priori that the observed chronic health conditions are a complete description of all health conditions that can lead to a work related health problem. (And indeed, the raw data have people who report a work disability while they do not report to have any of the observed health conditions.)

Respondent work limitations due to problems in domain *d* are given by:

$$Y_{ri}^{*}(d) = \beta(d)' X_i + \varepsilon_{ri}(d) \qquad (1.1)$$

For *d=4*, this equation applies to all respondents; for *d=1,2,3*, it only applies to those who report the corresponding health condition; for the others, $Y_{ri}*(d)$ will be minus infinity.

Response scales can vary with *d*. The response scale in domain *d* will be given by

$$\tau_i^{0}(d) = -\infty,\ \tau_i^{3}(d) = \infty,\ \tau_i^{1}(d) = \gamma^{1}(d)Vi;\ \tau_i^{2}(d) = \tau_i^{1}(d) + exp(\gamma^{2}(d)X_i).$$

Here we have merged the categories *moderately limited*, *severely limited* and *extremely limited/cannot work* into one category to reduce the total number of parameters to be estimated, reducing the five-point scale to a three-point scale. (The fourth model in Table 5 in the paper, 'Model combining moderate, severe, extreme,' shows that this reduction by itself has little impact on the main conclusions.)

If work limitations due to health problems in domain *d* were asked, they would be reported as

$$Y_{ri}(d) = j \text{ if } \tau_i^{j-1}(d) < Y_{rd}^{*} \leq \tau_i^{j}(d), \quad j=1,\ldots,3,\ d=1,\ldots,D\ (=4)$$

In the available data, however, work limitations in specific domains are not reported;[1] we only know whether there is *any* health problem that leads to work limitations. It seems reasonable to interpret this as the maximum of the work limitations in all domains:

$$Y_{ri} = \max\{Y_{ri}(1),\ldots,Y_{ri}(D)\}$$

To identify the model even in the standard case of no variation in response scales, some assumptions are needed to distinguish the four domains of reported overall work disability. The three domains *affect, pain* and *heart problems*, relate to reported health conditions: emotional problems (*has the doctor ever told you that you have emotional, nervous or psychiatric problems?*), pain (*do you often have pain?*) and heart problems (*has the doctor ever told you that you had a heart attack, coronary heart disease, angina, congestive heart failure, or other heart problems? Or has the doctor ever told you that you had a stroke or transient ischemic attack?*). As explained above, we assume that respondents can report a work disability[2] in one of the three domains *affect, pain*, or *CVD* only if they suffer from that health condition. The health condition dummies are included as covariates in our model (as in

[1] The HRS asks people reporting some work disability which domain(s) cause(s) the work disability. We do not use this information in the formal models, since the categories do not match our domains and since no such information is available for the Dutch data.

[2] That is, report a « yes » on the two-point scale or report a mild limitation or worse on the five point scale.

the benchmark model). Thus for someone who reports none of these three health conditions, only the domain *other* can lead to work disability. On the other hand, respondents who have an emotional problem but no heart condition and who do not suffer from pain, can be work disabled in either the *affect* domain or in the *other* domain (or both).

Moreover, we assume that work related health in the three domains *affect, pain* and *CVD* is not affected by other health conditions. Thus diabetes, cancer, lung disease, arthritis, or hypertension can only lead to work disability through the domain *other*. This implies zero restrictions on $\beta(1)$, $\beta(2)$ and $\beta(3)$.

Without vignettes, these assumptions are sufficient to identify $\beta(d)-\gamma^1(d)$, $d=1,...,4,$ but not the parameters of interest $\beta(d)$, $d=1,...,4$. Vignettes can be used to identify the parameters for the domains for which vignettes are available, $d=1,2,3$. No restrictions are imposed on $\gamma(1)$-$\gamma(3)$; these are identified by the vignettes in these domains (except for the three constant terms: as in the benchmark model, they are set to zero). Since here are no vignettes in the domain *other*, $\gamma^1(4)$ is not identified; only $\beta(4)-\gamma^1(4)$ is identified.

The vignette descriptions explicitly refer to problems in one domain, stating that the vignette-persons have no other health problems. Thus for the vignettes in domain d, it is reasonable to assume that work limitations in dimension d are larger than work limitations in other dimensions and completely determine the answer to the vignette work limitations question. This gives the following model for observed vignette evaluations, $Y_l(d)$, $l=1,...,L$ (L vignette descriptions for each dimension ; $L=5$ in our case), $d=1,...3$.

$$Y_{li}^*(d) = \theta_l(d)+\psi_l(d)\, Female_{li} + \varepsilon_{li}(d);$$

$$Y_{li}(d) = j \text{ if } \tau_i^{j-1}(d) < Y_{li}^*(d) \leq \tau_i^j(d), \quad j=1,2,3$$

$\varepsilon_{li}(d) \sim N(0,\sigma^2(d),$ independent of each other, of $\varepsilon_{ri}(d)$, and of X_i, V_i.

The vignette reports identify $\gamma^1(d)$ except for the constant terms $(d=1,2,3)$ and $\theta_l(d)$, $l=1,...,5$; $d=1,2,3,$ up to a constant term for each domain, and identify $\gamma^j(d)$ for $j>1$. The self-reports then identify $\beta(d)$. This is the same "correction" that was carried out in the benchmark ('one domain') model.

To estimate the model, an assumption needs to be made on the joint distribution of the errors. We assume joint normality and independence of each other and of the (thus exogenous) variables X_i.

The assumptions on the relation between the two-point scale and the five-point scale remain the same as before. We assume that this relation is the same for all domains.

Summarizing, we list the parameters in each equation of the multi-domain model:

- Respondent work disability in domains *d=1,2,3*: equation includes intercept, 5 demographics, and 6 interactions with the NL country dummy. The variance of the error term is normalized at 100 (fixing the scale). This gives 36 parameters.
- Response scales (2 thresholds) in domains *d=1,2,3*: demographics plus health conditions other than the one corresponding to this particular disability, with all interactions with dummy NL. To normalize location: no intercept in threshold 1. This gives 3*(23+24)=141 parameters.
- Respondent work disability in domain 4 *(other)*: intercept, 5 demographics, 4 health conditions, interactions with dummy NL. Error term has variance 100. This gives 20 parameters.
- Respondent work disability threshold 1 *other*: not identified.
- Respondent work disability threshold 2 *other*: only identified for NL (since there are no 5-point scale answers in the US on this domain, neither self-reports nor vignettes); 10 parameters.
- Vignette dummies, coefficients on gender of the vignette persons, standard deviations of vignettes. 3*(5+1+1)=21 parameters.
- Three auxiliary parameters transforming the five-point scale into the two-point scale (two standard deviations of idiosyncratic noise (independent across the two scales) and one for the weight of threshold 1; all assumed the same across domains and countries).
- In total: 36+147+20+10+21+3=237 parameters.[3]

Simulation Results

Table A3 presents simulation results based upon this model. The first panel gives the predictions for the age group 51-64 if everyone in each country uses their own response scales. For example, in The Netherlands, about 47 percent of those with an emotional

[3] To reduce the computational burden, we have not allowed for unobserved heterogeneity in the thresholds. The small difference between rows 2 and 3 in Table 5 in the paper suggest that this will not substantially affect the results.

condition classify themselves as work disabled because of this, versus only 27 percent in the US. The second panel shows that the difference is almost completely due to response scale differences: if the Dutch respondents would use the (higher) US response scales, then 27 percent of the Dutch with an emotional health condition would report themselves as disabled, the same as the predicted rate in the US.

Multiplying these numbers by the prevalence rates of emotional health problems (middle panel) gives work disability in the emotional health domain as a percentage of the total number of respondents in the age group. For the affect domain, once response scale differences are adjusted for, this is very similar in the two countries.

The results for work disability in the domains of pain and heart problems are quite different. The prevalence rate for pain is smaller in the US than in The Netherlands, as we saw before. In The Netherlands, the probability that people who often have pain would report a pain related work disability is almost twice as large as in the US. While the difference would be a lot smaller if the Dutch would use the US response scales, it would not disappear. Even then, work disability in the pain domain would explain a more than 17 percent work disability rate in The Netherlands compared to 10 percent in the US.

For heart problems, the response scales in the two countries are rather similar, so that there is only a small adjustment if response scale differences are controlled for. US respondents more often report that the doctor has told them that they have a heart problem than Dutch respondents, but Dutch respondents with heart problems have a substantially larger probability to be work disabled. Since the latter difference is larger than the former, the rate of heart problems related work disability is somewhat larger in The Netherlands than in the US.

Comparing the three domains, we find that there is more pain related work disability than affect or CVD related work disability in both countries. For the US, this is at least qualitatively most in line with the HRS data on the most important source of work disability – back, neck and spine problems.

The fourth row in each panel shows how many respondents suffer from work disability in at least one of the three domains. In the US, these three domains give a work disability rate of 15.8 percent, 69 percent of the total work disability rate in this age group. In The Netherlands, and using Dutch response scales, the three domains explain

almost 78 percent of total work disability. Combining the three domains, the difference in work disability in either of these three domains between The Netherlands and the US reduces from 13.2 percent to 6.9 percent if response scale differences are adjusted for. Thus about half of the gap is due to response scale differences, a conclusion similar to that based upon the benchmark model.

Table A3 Predicted Work Disability Age Group 51-64 – US versus NL

Panel 1- Predictions using own response scales:

Domain	Work disability in group with health condition		Prevalence of health condition		Work disability in population	
	NL	US	NL	US	NL	US
affect	47.2	27.1	10.5	14.3	5.0	3.9
pain	65.6	36.1	33.7	27.6	22.1	10.0
cvd	41.7	26.7	12.2	15.6	5.1	4.2
a,p,c	62.9	37.7	46.1	42.0	29.0	15.8
other					12.4	9.4
total					37.2	23.1

Panel 2-Predictions using US response scales

Domain	Work disability in group with health condition		Prevalence of health condition		Work disability in population	
	NL	US	NL	US	NL	US
affect	27.3	27.1	10.5	14.3	2.9	3.9
pain	51.5	36.1	33.7	27.6	17.4	10.0
cvd	37.1	26.7	12.2	15.6	4.5	4.2
a,p,c	49.2	37.7	46.1	42.0	22.7	15.8
other					8.3	9.4
total					28.5	23.1

Notes: CentER Savings Survey 2003 for The Netherlands and HRS 1998 for the US, weighted with sampling weights.

Work disability rates due to other health problems than heart problems, emotional problems, or pain, are 12.4 percent in The Netherlands with Dutch response scales, and 9.4 percent in the US with US response scales, and part of the difference can be due to different response scales. Combining *other* with the three domains affect, pain, and CVD gives total work disability rates using country specific response scales of 37.2 percent and 23.1 percent, close to the work disability rates in the raw data. For this domain, we cannot correct for response scale differences in the same way as for the other three domains, since no vignettes on *other* are available. In the second panel of Table 9, we have assumed that DIF for *other* is a weighted mean of DIF for *affect, pain,* and *cvd*, with weights determined by the contribution of these three to explaining work disability, as given in the final columns of Table A3. Estimated total work disability in The Netherlands on US response scales would then be 28.5 percent (total-2 in the bottom panel). Adjusting for response scale differences reduces the cross-country difference in overall work disability from 14.2 percent-points to 5.4 percent-points. This is quite similar to the conclusion obtained with the benchmark model.